WATER RESOURCES RESEARCH, VOL. 37, NO. 10, PAGES 2521–2535, OCTOBER 2001

Bayesian recursive parameter estimation for hydrologic models

M. Thiemann,[1] M. Trosset,[2] H. Gupta, and S. Sorooshian

Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona

Abstract. The uncertainty in a given hydrologic prediction is the compound effect of the parameter, data, and structural uncertainties associated with the underlying model. In general, therefore, the confidence in a hydrologic prediction can be improved by reducing the uncertainty associated with the parameter estimates. However, the classical approach to doing this via model calibration typically requires that considerable amounts of data be collected and assimilated before the model can be used. This limitation becomes immediately apparent when hydrologic predictions must be generated for a previously ungauged watershed that has only recently been instrumented. This paper presents the framework for a Bayesian recursive estimation approach to hydrologic prediction that can be used for simultaneous parameter estimation and prediction in an operational setting. The prediction is described in terms of the probabilities associated with different output values. The uncertainty associated with the parameter estimates is updated (reduced) recursively, resulting in smaller prediction uncertainties as measurement data are successively assimilated. The effectiveness and efficiency of the method are illustrated in the context of two models: a simple unit hydrograph model and the more complex Sacramento soil moisture accounting model, using data from the Leaf River basin in Mississippi.

1. Introduction and Scope

The science of hydrology is occupied primarily with the processes that make up the hydrologic water cycle and the hydrologic implications of climatic and anthropogenic changes. A variety of hydrologic models have been developed to facilitate this task, i.e., to simulate the water cycle (or a portion of it) for a region of interest. While some of the parameters of these models may represent measurable attributes such as watershed area, others typically represent conceptual (effective) attributes such as mean hydraulic conductivity or rates of drainage for hypothetical lumped water storages. Therefore the numerical values of many of the parameters are not easily inferred from quantities that can be measured and must be specified by calibration.

The aim of model calibration procedures is to reduce the uncertainty in the correct choice of the parameter values (parameter uncertainty) while accounting for uncertainties in the values of the measured input-output time series (data uncertainty) and uncertainties in the structural ability of the model to simulate the processes of interest (model uncertainty or structural uncertainty). The collective impact of the parameter, data, and structural uncertainties gives rise to uncertainties in the predictions made using the model (prediction uncertainty).

During the past two decades, much work has been done to understand and improve hydrologic model calibration methods. Sophisticated measures have been developed to measure the closeness of fit between the predicted and observed outputs [e.g., *Sorooshian and Dracup*, 1980; *Sorooshian*, 1981], effective and efficient global optimization strategies such as the shuffled complex evolution developed by the University of Arizona (SCE-UA) algorithm have been developed to search the parameter space [*Duan et al.*, 1992, 1993; *Sorooshian et al.*, 1993], and procedures have been developed for the statistical analysis of parameter uncertainty [e.g., *Spear and Hornberger*, 1980; *Jones*, 1983; *Kuczera*, 1988; *Spear et al.*, 1994]. Recent innovations include generalized likelihood uncertainty methods [*Binley and Beven*, 1991; *Romanowicz et al.*, 1994; *Franks and Beven*, 1997] and multicriteria methods [*Gupta et al.*, 1998; *Yapo et al.*, 1998; *Gupta et al.*, 1999; *Bastidas et al.*, 1999].

The primary focus of the aforementioned research has been the "batch" calibration approach. This approach assumes the model parameter values to be time invariant and typically requires that considerable data be collected before the procedure can be implemented to arrive at "optimal" estimates for those parameters. Only after the model has been calibrated and tested can it be employed to make predictions. Further, by focusing on an optimal parameter set the approach implicitly ignores parameter uncertainty. Calibrated models are typically used to generate model predictions, which are then reported without estimates of the underlying prediction uncertainty. Only recently has attention been given to methods that incorporate prediction uncertainty [e.g., *Beven and Binley*, 1992; *Kuczera et al.*, 1993; *Mroczkowski et al.*, 1997; *Franks and Beven*, 1997].

The limitations of the batch calibration approach become immediately apparent when we attempt to generate hydrologic predictions for a previously ungauged watershed that has only recently been instrumented. The lack of sufficient historical data makes it impossible to apply batch methods to calibrate a model for the watershed; for example, calibration of the Sacramento soil moisture accounting (SAC-SMA) model [*Burnash et al.*, 1973] used by the U.S. National Weather Service for flood forecasting typically requires at least 11 years of data.

This paper develops the framework for a recursive approach to hydrologic prediction that performs parameter estimation and output prediction in "on-line" mode. The method, called

[1]Now at Riverside Technology, Inc., Fort Collins, Colorado.

[2]Now at Department of Mathematics, College of William and Mary, Williamsburg, Virginia.

Paper number 2000WR900405.
0043-1397/01/2000WR900405$09.00

Bayesian Recursive Estimation (BaRE), requires only an initial guess of the region of the parameter estimates to be specified before the model can be used to begin the generation of one-step-ahead (and multiple-step-ahead) predictions. These predictions are described in terms of the probabilities associated with different output values (or can be summarized in terms of a "most likely" prediction and a "Bayesian confidence interval"). As might be expected, the uncertainty associated with the prediction will be relatively large in the beginning. A recursive procedure is used to update (reduce) the uncertainty associated with the parameter estimates as successive input-output measurement data are assimilated. The reduced parameter uncertainty results in smaller prediction uncertainties.

This paper is organized as follows. Section 2 presents a review of the mathematical basis for Bayesian inference, its relationship to prediction and calibration, its role in recursive inference, and some comments about the Generalized Likelihood Uncertainty Estimation (GLUE) procedure presented by *Beven and Binley* [1992]. The BaRE methodology and algorithmic procedure is developed in section 3, while section 4 illustrates the usefulness of BaRE methodology by means of two preliminary case studies involving streamflow prediction. Finally, section 5 discusses the strengths and weaknesses of the procedure and makes suggestions for further improvements.

2. Bayesian Inference

The fundamental problem with which we are concerned is to predict (and to quantify the uncertainty in our predictions of) measurable hydrologic quantities (outputs) from observed hydrologic data (inputs) using a specified family of mathematical models that simulate actual input-output relations. For example, suppose that we are given the following: (1) a family of mathematical models that simulate the runoff response to rainfall for a specified watershed, perhaps together with some information about which models in the family are most plausible and (2) time series of observed input-output data (e.g., precipitation, potential evapotranspiration, and streamflow) for the specified watershed from some time $t = 1$ in the past to the present time T. Then a typical problem might be how to predict streamflow at time $T + 1$ and further into the future.

This section is organized as follows. In section 2.1 we formulate problems like the above as problems of Bayesian inference for nonlinear regression. We explain why the Bayesian framework for statistical inference is admirably suited to such problems and assemble the basic mathematical concepts which will be employed. Our notation is intended to facilitate comparison with the comprehensive exposition of Bayesian methodology in the classic treatise by *Box and Tiao* [1973]. In section 2.2 we show how the Bayesian formulation permits hydrologists to quantify uncertainty about prediction in a natural and meaningful way. This is accomplished without recourse to calibration, which can be conceptually and/or practically problematic for certain applications. However, because hydrologists may encounter applications for which calibration is meaningful, in section 2.3 we explain how to calibrate within the Bayesian framework. We introduce a recursive formulation of the Bayesian framework in section 2.4. Our application of Bayesian methodology to problems in hydrology has certain conceptual similarities to the GLUE procedure of *Beven and Binley* [1992] but differs from it in certain important respects that we explore in section 2.5.

2.1. Formulation

We are interested in mathematical models that predict outputs from inputs. The models are indexed by parameters, which may (or may not) be physically interpretable. Following *Box and Tiao* [1973], we denote the predictive model by η, the vector of all its parameters by θ, inputs by ξ, and outputs by y. Given model parameter values of θ and input values of ξ, we predict output values

$$\hat{y} = \eta(\xi|\theta). \tag{1}$$

Given observed output values of y, we denote the residual errors from our predictions by

$$\varepsilon = y - \hat{y} \tag{2}$$

and summarize these relations by writing

$$y = \eta(\xi|\theta) + \varepsilon. \tag{3}$$

This is a standard formulation of nonlinear regression.

Next we elaborate on the structure of the above quantities. First, we assume that

$$\theta \in \Theta \subseteq \mathcal{R}^k, \tag{4}$$

where $\mathcal{R}^k$ denotes k-dimensional Euclidean space.

Assume that there are p inputs, indexed by $r = 1, \ldots, p$ and s outputs indexed by $s = 1, \ldots, m$, available for time steps $t = 1, \ldots, T$. Thus ξ_t^r denotes the value of input r, measured at time t, and y_t^s denotes the value of output s, measured at time t. Often, we will manipulate entire arrays of these quantities, e.g., writing $\boldsymbol{\xi}$ and $\mathbf{y}$ to denote all of the inputs and outputs up to the current time T. On this basis we wish to predict the as yet unobserved outputs for the following time steps. For simplicity, we restrict this study to the next time step $T + 1$. Accordingly, the nonlinear regression equation of interest is

$$y_{T+1} = \eta(\boldsymbol{\xi}|\theta) + \varepsilon_{T+1}. \tag{5}$$

Equation (5) contains various sources of potential prediction error: (1) measurement error in which, for example, it may be that $\eta(\ \ |\theta_1)$ is an accurate representation of the physical data-generating process but that the devices used to measure y_{T+1} are somewhat inaccurate; (2) parameter identification error in which it may be that $\eta(\ \ |\theta_1)$ is not an accurate representation of the physical data-generating process but that $\eta(\ \ |\theta_2)$ is accurate; and (3) model specification error in which it may be that for no $\theta \in \Theta$ is the model $\eta(\ \ |\theta)$ an accurate representation of the physical data-generating process. Thus, regardless of the sophistication of the mathematical model η, there remains uncertainty about the predictions that we use η to make. The purpose of this paper is to explicate a formal methodology for quantifying that uncertainty.

The uncertainties which we have identified so far differ in a fundamental respect. Uncertainty about the measurements can be formalized in a familiar way: The residuals are assumed to be drawn from a suitable probability distribution. In contrast, uncertainty about the parameters (about which we will say much) and uncertainty about the model itself (about which we will say little) express degrees of belief. However, one way to quantify the latter type of uncertainty is to introduce suitable subjective probability distributions. Expressing uncertainty in terms of probabilities is one of the defining features of Bayesian inference. Once this has been done in an appropriate way,

inference becomes a matter of manipulating various probability distributions according to the rules of mathematical probability.

We begin by introducing a probability distribution on the possible parameter sets and denoting its density function by $p(\theta)$. This is an unconditional distribution in the sense that it does not depend on the currently available values of the inputs ξ or outputs $\mathbf{y}$. In the context of Bayesian inference it is usually called the prior distribution because it can be specified before any data are collected. The purpose of the prior distribution is to quantify initial uncertainty about which parameter values should be employed. Usually, the prior distribution expresses a great deal of initial uncertainty. For example, *Beven and Binley* [1992] suggested restricting attention to a large rectangle of parameter values and imposing a uniform prior distribution on that rectangle. However, the prior distribution can also reflect knowledge about the parameter values based on the analysis of historical data from previous events or from another system (e.g., watershed) having similar characteristics.

All of the other probability distributions with which we will be concerned are conditional distributions. We will denote their density functions by writing expressions of the form $p(\ |\ )$. Arguments on the left of the vertical bar denote the variables of the density; arguments on the right of the vertical bar denote the fixed values on which the density is conditioned. Context is to be inferred by examining which symbols appear in which roles.

Thus, $p(y_{T+1}, \mathbf{y}|\xi; \theta)$ is the conditional probability density of the outputs y_0 to y_{T+1}, given the inputs and the model parameters. The functional form of this density must be specified by the hydrologist, usually by selecting a plausible family of distributions for the residual errors in (5). We will have more to say about how this might be done in section 3.

Once the prior density and conditional density have been specified, everything else follows automatically. The conditional density of the outputs and the model parameters, given only the inputs, is

$$p(y_{T+1}, \mathbf{y};\ \theta|\xi) = p(y_{T+1}, \mathbf{y}|\xi;\ \theta)p(\theta), \tag{6}$$

and the conditional density of those outputs to be predicted and of the model parameters, given all of the data that have been collected, is

$$p(y_{T+1};\ \theta|\xi, \mathbf{y}) = Cp(y_{T+1}, \mathbf{y};\ \theta|\xi), \tag{7}$$

where $C^{-1} = \int p(y_{T+1}, \mathbf{y};\ \theta|\xi)d(\mathbf{y})$. Therefore (7) gives a posterior density that quantifies all the uncertainty that remains about the outputs to be predicted and the model parameters after the information in the data has been assimilated. The derivation of this density is usually identified as Bayes theorem.

What we do next depends on whether we are interested in quantifying the uncertainty in our predictions or in selecting specific values for the parameters of the model (i.e., model calibration). Notice that (7) is a joint posterior density of the outputs to be predicted and of the model parameters to be estimated. If we are interested in prediction, then we require the marginal posterior density of the outputs to be predicted; if we are interested in calibration, then we require the marginal posterior density of the model parameters to be estimated. Of course, one obtains a marginal density from a joint density by integrating the joint density with respect to the variables that are not of interest. Thus, in the Bayesian formulation of the problem, prediction and calibration are accomplished by symmetric operations. We proceed to describe these operations in greater detail.

2.2. Prediction

Prediction is the problem of describing a plausible set of values for y_{T+1}, the as yet unobserved outputs. To predict, we first compute the marginal posterior density of y_{T+1}:

$$p(y_{T+1}|\xi, \mathbf{y}) = \int_\Theta p(y_{T+1};\ \theta|\xi, \mathbf{y})\ d\theta. \tag{8}$$

Once this density has been computed, prediction is merely a question of identifying and reporting meaningful summaries of it. For example, one might use one of the following methods.

1. Report a single-predicted value of y_{T+1} by stating some measure of central tendency of (8), e.g., its mean, median, or mode.
2. Report a set of plausible values of y_{T+1} by constructing a region of highest posterior density (HPD) for (8). A subset R of the domain of p is called an HPD region of content $1 - \alpha$ if $P(R) = 1 - \alpha$ and $p(y_1) \geq p(y_2)$ for any $y_1 \in R$ and $y_2 \notin R$.
3. If there is only $m = 1$ output to predict, then one can summarize (8) simply by reporting several percentiles of it. If there are $m \geq 2$ outputs to predict, then construction of an HPD region may be time consuming. If the outputs are not highly correlated, then one may prefer to calculate marginal percentiles for each output.

2.3. Calibration

Calibration is the problem of describing a plausible set of values for θ, the parameters of the models. To calibrate, we must first compute the marginal posterior density of θ:

$$p(\theta|\xi, \mathbf{y}) = \int p(y_{T+1};\ \theta|\xi, \mathbf{y})\ dy_{T+1}. \tag{9}$$

An HPD region for this density describes a set of plausible models.

If we desire a formal estimate of θ, then we must specify a loss function. Let $L(\theta,a)$ denote the loss incurred by the hydrologist if $\theta \in \Theta$ is true and $a \in \Theta$ is estimated. The Bayes estimate of θ is the $a \in \Theta$ that minimizes the function

$$f(a) = \int_\Theta L(\theta, a)p(\theta|\xi, \mathbf{y})\ d\theta. \tag{10}$$

If $L(\theta,a) = (a - \theta)^2$, then the Bayes estimate is the mean of (9).

Of course, $L(\theta,a) = (a - \theta)^2$ is rarely an appropriate loss function for calibrating hydrologic models. The parameters may lack intrinsic meaning, in which case measuring errors by imposing a metric on the parameter space seems contrived. Furthermore, because the HPD region may be disconnected, several sets of parameter values are plausible but their mean is not. In view of these considerations, *Beven and Binley* [1992] persuasively argued that losses should be measured in the space of possible outputs, not in the space of possible parameter values. To do so, let

$$L(\theta, a) = \int \chi[\eta(\xi|\theta),\ \eta(\xi|a)]w(d\xi), \tag{11}$$

where χ measures the badness of fit between two predicted series of outputs and w is a probability distribution that weights the possible inputs.

Furthermore, we note that we could define several loss functions and construct an estimate by multiobjective optimization [*Gupta et al.*, 1998; *Yapo et al.*, 1998]. In this paper, we assume that all objectives of interest have been collected in a single loss function.

2.4. Recursive Inference

Bayesian inference provides a formal mechanism for combining previous information with new information. In the words of *Box and Tiao* [1973, p. 12], "Bayes' theorem describes, in a fundamental way, the process of learning from experience, and shows how knowledge about the states of nature represented by θ is continually modified as new data become available." In the present context, the Bayes theorem provides a way of updating information about the hydrologic model as more data are collected.

Suppose we have collected input-output data ($\boldsymbol{\xi}$,$\mathbf{y}$) from time $t = 1$ to the current time $t = T$. Our current knowledge about the model parameters θ is contained in $p(\theta|\boldsymbol{\xi}, \mathbf{y})$, the marginal posterior of θ defined in (9). Observing that

$$p(\theta|\boldsymbol{\xi}, \mathbf{y})p(\mathbf{y}) = p(\mathbf{y}; \theta|\boldsymbol{\xi}) = p(\mathbf{y}|\boldsymbol{\xi}; \theta)p(\theta), \quad (12)$$

we see that

$$p(\theta|\boldsymbol{\xi}, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\xi}; \theta)p(\theta). \quad (13)$$

Now suppose that we collect additional input-output data $(\xi_{T+1}; y_{T+1})$. By the same reasoning,

$$p(\theta|\xi_{T+1}, \boldsymbol{\xi}, y_{T+1}\mathbf{y}) \propto p(y_{T+1}, \mathbf{y}|\xi_{T+1}, \boldsymbol{\xi}; \theta)p(\theta). \quad (14)$$

If y_{T+1} and $\mathbf{y}$ are conditionally independent given $(\xi_{T+1}, \boldsymbol{\xi})$, then

$$\begin{aligned} p(y_{T+1}, \mathbf{y}|\xi_{T+1}, \boldsymbol{\xi}; \theta)p(\theta) &= p(y_{T+1}|\xi_{T+1}, \boldsymbol{\xi}; \theta)p(\mathbf{y}|\boldsymbol{\xi}; \theta)p(\theta) \\ &\propto p(y_{T+1}|\xi_{T+1}, \boldsymbol{\xi}; \theta)p(\theta|\boldsymbol{\xi}, \mathbf{y}). \end{aligned}$$

Thus, (14) becomes

$$p(\theta|\xi_{T+1}, \boldsymbol{\xi}, y_{T+1}, \mathbf{y}) \propto p(y_{T+1}|\xi_{T+1}, \boldsymbol{\xi}; \theta)p(\theta|\boldsymbol{\xi}, \mathbf{y}), \quad (15)$$

a recursive formula for updating information about θ. Notice that the marginal posterior density in (9) plays the same role in (15) that the prior density played in (6). Finally, notice that we can substitute (15) into (8).

2.5. Bayesian Inference and GLUE

The sequential nature of Bayes theorem follows from elementary properties of joint, marginal, and conditional probability density functions. The Bayesian paradigm, in which densities are manipulated according to the rules of mathematical probability, has many appealing consequences. For example, it guarantees that the order in which n mutually independent previous events occurred does not affect inferences about the current event. However, if the probability density functions in (13) and (15) are replaced by other quantities, then these relations cease to be meaningful updating formulae.

Such a substitution (in a somewhat less general framework) is a distinctive feature of the GLUE procedure of *Beven and Binley* [1992]. If we write $p(y_{T+1}, \mathbf{y}|\boldsymbol{\xi}; \theta)$ as a function of its unknown arguments

$$l(y_{T+1}; \theta|\boldsymbol{\xi}, \mathbf{y}) = p(y_{T+1}, \mathbf{y}|\boldsymbol{\xi}; \theta), \quad (16)$$

then l is traditionally called the likelihood function of the unknown quantities given the known quantities [*Fisher*, 1922]. *Beven and Binley* [1992] replaced the likelihood function l with a "likelihood measure" L, which they take to be a measure of how well the model predictions fit the observations [*Beven and Binley*, 1992, p. 281]: "We use the term likelihood here in a very general sense, as a fuzzy belief, or probabilistic measure of how well the model conforms to the observed behavior of the system, and not in the restricted sense of maximum likelihood theory...."

The GLUE methodology does not respect the rules of Bayesian inference. Formally, GLUE admits any likelihood measure L in place of the likelihood function l. Some likelihood measures use a "likelihood shape factor" N, which is set by the user to control the peakedness of the likelihood function [*Binley and Beven*, 1991; *Freer et al.*, 1996; *Franks and Beven*, 1997; *Franks et al.*, 1998]. In addition, a user-defined "behavioral threshold" T that separates multiple "behavioral" (accurate) from "nonbehavioral" (less accurate) simulations is commonly employed [*Beven and Binley*, 1992; *Freer et al.*, 1996; *Franks and Beven*, 1997; *Franks et al.*, 1998]. Both N and T demonstrate the emphasis on the use of subjective, user-defined parameters within the GLUE methodology to evaluate parameter probabilities. The use of N and T also supports the philosophy behind GLUE that there exists no optimal model due to parameter insensitivity, parameter interactions, and nonuniqueness of model structures. This is termed as "equifinality" by *Beven* [1993].

In contrast, the BaRE methodology draws inferences within the Bayesian framework. Respecting this framework leads to the examination of GLUE's likelihood measures from several new perspectives.

1. The examples of likelihood measures discussed by *Beven and Binley* [1992] are measures of fit. They are naturally accommodated by the Bayesian framework as possible choices of χ in (11).
2. *Beven and Binley* [1992] chose L to be the reciprocal of the sum of the squared residuals raised to a power specified by the user. In section 3.2 we show that if the measurement errors have an exponential power distribution, then a certain power of the sum of the squared residuals is actually proportional to the posterior distribution of θ.
3. *Beven and Binley* [1992, p. 287] regarded L as "a fuzzy measure that reflects the degree of belief of the modeler." In the Bayesian framework such information is incorporated into the inference via the prior distribution of θ. When the prior distribution is derived from the sample, it is sometimes called an empirical prior distribution. Accordingly, we note the possibility of using L as an empirical prior distribution for θ.

3. Bayesian Recursive Estimation (BaRE) Algorithm

3.1. Selection of the Error Model

Implementation of the Bayesian inference scheme requires selection of a functional form for the conditional probability density of the outputs by selecting a plausible family of distributions for the residual errors in (5). In this paper, we consider

the case of a single input ($p = 1$) and a single output ($m = 1$). We shall assume that there exists a one-to-one and continuous invertible transformation

$$z = g(y), \tag{17}$$

such that the measurement errors in the transformed output space, given by

$$v = g(y) - g(\hat{y}) \tag{18}$$

are mutually independent, each having the exponential power density $E(\sigma,\beta)$ described by *Box and Tiao* [1973, section 3.5]:

$$p(v|\sigma, \beta) = \omega(\beta)\sigma^{-1} \exp\left[-c(\beta)|v/\sigma|^{2/(1+\beta)}\right], \tag{19}$$

where

$$c(\beta) = \left\{\frac{\Gamma[3(1+\beta)/2]}{\Gamma[(1+\beta)/2]}\right\}^{1/(1+\beta)},$$

$$\omega(\beta) = \frac{\{\Gamma[3(1+\beta)/2]\}^{1/2}}{(1+\beta)\{\Gamma[(1+\beta)/2]\}^{3/2}}, \tag{20}$$

the "shape parameter" $\beta \in (-1,1]$ is fixed, and the standard deviation of the measurement errors $\sigma > 0$ is unknown but constant with respect to time. The parameter β can be regarded as a measure of kurtosis, indicating the extent of the "nonnormality" of the parent population; for example, the density is normal when $\beta = 0$, double exponential when $\beta = 1$, and tends to a uniform distribution as $\beta \to -1$. Note that the transformation g allows us to handle the nonconstant variance (heteroscedastic) error situations which are common in hydrology, that is, where the variance of the measurement error associated with y varies with flow level [e.g., see *Sorooshian and Dracup*, 1980]. Although not explicitly handled in this paper, the error model could also be extended to autocorrelated errors [e.g., see *Sorooshian and Dracup*, 1980; *Kuczera*, 1988] for data which indicate that the error series of the used model is, in fact, not independent.

3.2. Derivation of the Recursive Form

Given the error model specified by (20), the conditional density of the transformed outputs observed at times $t = 1, \ldots, T$ is

$$p(\mathbf{z}|\boldsymbol{\xi}; \theta, \sigma, \beta) = \prod_{t=1}^{T} \frac{\omega(\beta)}{\sigma} \exp\left[-c(\beta)\left|\frac{z_t - g[\eta(\boldsymbol{\xi}|\theta)]}{\sigma}\right|^{2/(1+\beta)}\right] \tag{21}$$

$$p(\mathbf{z}|\boldsymbol{\xi};\ \theta,\ \sigma,\ \beta) = \left[\frac{\omega(\beta)}{\sigma}\right]^{T} \exp\left[-c(\beta)\sum_{t=1}^{T}\left|\frac{v_t(\theta)}{\sigma}\right|^{2/(1+\beta)}\right]. \tag{22}$$

Box and Tiao [1973] treated the standard deviation parameter σ as a "nuisance" parameter and showed that by assuming a noninformative prior density of the form $p(\theta,\sigma|\beta) \propto \sigma^{-1}$, the influence of σ can be integrated out, leading to the following form for the posterior density for θ:

$$p(\theta|\boldsymbol{\xi}, \mathbf{z}; \beta) \propto [M(\theta)]^{-(T)(1+\beta)/2}, \tag{23}$$

where

$$M(\theta) = \sum_{t=1}^{T} |z_t - g[\eta(\boldsymbol{\xi}|\theta)]|^{2/(1+\beta)} = \sum_{t=1}^{T} |v_t(\theta)|^{2/(1+\beta)}. \tag{24}$$

Notice that in the case of normally distributed errors ($\beta = 0$), M is simply the familiar "sum of the squared measurement errors" function that is commonly minimized in model calibration. Note also that as $T \to \infty$, this distribution begins to concentrate on the global minimizer of $M(\theta)$, i.e., on θ^*. Thus our calibration procedure is a consistent estimator of θ.

Unfortunately, the use of (24) for computing the posterior density of the parameters given the data leads to numerical stability problems as $T \to \infty$; $M(\theta)$ tends to a constant value when evaluated at a given θ, but the term $[M(\theta)]^{-(T)(1+\beta)/2}$ grows without bound. Further, there is a practical advantage to computing an explicit estimate for the unknown parameter σ. Therefore we follow *Box and Tiao* [1973] in assuming that θ and $\log \sigma$ are independent, but instead of assuming a locally uniform prior in θ and σ and integrating out the influence of σ, we compute the maximum likelihood estimate of σ_t at the current time step by maximizing (23) with respect to σ:

$$\hat{\sigma}_T(\theta)^{2/(1+\beta)} = \frac{1}{T}\frac{2c(\beta)}{(1+\beta)}\sum_{t=1}^{T}|v_t(\theta)|^{2/(1+\beta)}. \tag{25}$$

Numerical experiments have shown that the practical implications of this difference in strategy are insignificant. Equation (25) can be rewritten in recursive form as

$$\hat{\sigma}_T(\theta)^{2/(1+\beta)} = \frac{T-1}{T}\hat{\sigma}_{T-1}(\theta)^{2/(1+\beta)} + \frac{1}{T}\frac{2c(\beta)}{(1+\beta)}|v_T(\theta)|^{2/(1+\beta)}. \tag{26}$$

Substitution of (26) into (19) leads to the following recursive formulation for estimating the posterior density for θ.

$$p(\theta|\boldsymbol{\xi}, z_{T+1}, \mathbf{z}; \beta) \propto N_T(\theta)p(\theta|\boldsymbol{\xi}, \mathbf{z}; \beta), \tag{27}$$

where

$$N_T(\theta) = \frac{1}{\hat{\sigma}_T(\theta)}\exp\left[-c(\beta)\left|\frac{v_T(\theta)}{\hat{\sigma}_T(\theta)}\right|^{2/(1+\beta)}\right]. \tag{28}$$

As more and more data are collected, the posterior density described above will tend to concentrate on the set of values of θ that minimize $M(\theta)$. Of course, this set of minimizing values will be sensitive to the choice of the predictive model, the error model and transformation, and the data used. In addition, the "best" parameter sets can vary over time, as $v_T(\theta)$ in (28) will always emphasize the ability of a certain parameter set to reproduce the most recent observation. If a formerly superior parameter set cannot simulate the desired processes at the current time T, it will receive lower weight in the computation of the posterior density, hence allowing a shift in the estimated elements of θ^*.

3.3. BaRE Algorithm

In view of the inevitably complicated nature of the hydrologic model η, it is evident that an explicit expression for the various posterior and conditional densities is not possible. Instead, we employ the power of digital computing to construct approximations of these various quantities via Monte Carlo simulation as described in the following algorithm:

3.3.1. Preparation. Preparation for use of the BaRE algorithm involves the following steps:

1. Choose the system model $y = \eta(\theta|\boldsymbol{\xi})$, the transformation model $z = g(y)$, and the error model $v \sim E(\sigma,\beta)$.
2. Select a value for the kurtosis parameter β and an initial estimate for the variance parameter $\hat{\sigma}_0$ of the error model.

3. Define the feasible parameter space Θ in terms of upper and lower limits for each element of θ.

4. Define the prior probability distribution $p_0(\theta)$ for θ. In the absence of any information, select a uniform prior.

3.3.2. Discrete sampling. Sample n distinct parameter sets θ^i, $i = 1, \ldots, n$ from a uniform distribution on Θ. Larger values for n will result in more accurate approximations for the various density functions.

3.3.3. Initialization. To establish initialization, perform the following:

1. Set $T = 0$.

2. Initialize the prior $p(\theta^i|\xi, \mathbf{z}; \beta) = p_0(\theta)$ and $\hat{\sigma}_0(\theta^i) = \hat{\sigma}_0$ for each parameter set ($i = 1, \ldots, n$).

3.3.4. Prediction of the output. Here we estimate the output prediction uncertainty associated only with the current uncertainty in the model parameters. If the model has no structural error, this is an estimate of the uncertainty in prediction of the "true" output, i.e., of observations that are not disturbed by measurement errors.

1. Compute the transformed model output $\hat{z}_{T+1}(\theta^i) = g(\eta(\theta^i|\xi))$ for each parameter set ($i = 1, \ldots, n$).

2. Sort the outputs in order of increasing magnitude in preparation for the next step (step 3); that is, compute $\{j = 1, \ldots, n\} = sort\{i = 1, \ldots, n\}$ such that $\hat{z}_{T+1}(\theta^j) \geq \hat{z}_{T+1}(\theta^{j-1})$.

3. Compute the cumulative distribution function of the predicted output in the transformed space $P(z_{T+1} \leq a|\xi, \mathbf{z}, \beta) = \sum_{j=1}^{k} p(\theta^i|\xi, \mathbf{z}, \beta)$, where $\hat{z}_{T+1}(\theta^k) \leq a < \hat{z}_{T+1}(\theta^{k+1})$.

4. Compute appropriate percentiles to define the HPD region for $\hat{z}_{T+1}$, and transform these to the original output space.

3.3.5. Prediction of the output measurement. Here we increase the prediction uncertainty to account for the total residual error (i.e., structural error plus as yet unobserved output measurement error) as estimated by the error model.

1. Define the output region of interest. (1) Find indices $[l, u]$ such that $\hat{z}_{T+1}(\theta^l) = \min\{\hat{z}_{T+1}(\theta^i), i = 1, \ldots, n\}$ and $\hat{z}_{T+1}(\theta^u) = max\{\hat{z}_{T+1}(\theta^i), i = 1, \ldots, n\}$. (2) Define the upper and lower output values $[a_{T+1}^{\min}, a_{T+1}^{\max}]$ such that $a_{T+1}^{\min} = \hat{z}_{T+1}(\theta^l) - 2\hat{\sigma}_T(\theta^l)$ and $a_{T+1}^{\max} = \hat{z}_{T+1}(\theta^u) + 2\hat{\sigma}_T(\theta^u)$. (3) Define a discrete set of n_a uniformly spaced sampling points b on the output range (e.g., $n_a = 100$) according to

$$b = \left\{ a_{T+1}^{\min} + \frac{a_{T+1}^{\max} - a_{T+1}^{\min}}{n_a - 1}(m-1), m = 1, \ldots, n_a \right\}.$$

2. Compute the probability density of the as yet unobserved output measurement in the transformed space

$$p(\tilde{z}_{T+1} = b_k|\xi, \mathbf{y}) = C \sum_{i=1}^{n} N_{T+1}(\theta|\xi, \hat{\sigma}_T, \beta) \Big|_{z_{T+1}=b_k} p(\theta^i|\xi, \mathbf{y}),$$

where C is a constant that normalizes the total probability mass to 1.0.

3. Compute the cumulative density of the as yet unobserved output measurement in the transformed space

$$P(\tilde{z}_{T+1} \leq b_k|\xi, \mathbf{y}) = \sum_{j=1}^{k} p(\tilde{z}_{T+1} = b_k|\xi, \mathbf{y}).$$

4. Compute appropriate percentiles to define the HPD region for $\tilde{z}_{T+1}$ and transform these to the original output space.

3.3.6. Updating. When the observation of y_{T+1} becomes available, the following steps are taken.

1. Compute the transformed measurement $z_{T+1} = g(y_{T+1})$.

2. Update the estimates of the error model variance $\hat{\sigma}_{T+1}(\theta^i)$ for $i = 1, \ldots, n$ according to (26).

3. Compute the posterior parameter density

$$p(\theta^i|\xi, \mathbf{z}, z_{T+1}) = CN_{T+1}(\theta|\xi, \hat{\sigma}_{T+1}, \beta)|_{z_{T+1}} p(\theta^i|\xi, \mathbf{y})$$

for $i = 1, \ldots, n$ and use as the prior distribution for the next time step.

4. Set $T = T + 1$ and resume with prediction of the output.

4. Case Studies

We illustrate the power and applicability of the Bayesian Recursive Estimation scheme by means of two simple case studies. The first is a synthetic study using the two-parameter Nash- cascade model [*Nash*, 1960]. This illustrates the ability of the BaRE method to locate (i.e., assign) the highest probability to the region of the known true parameter values. The second study explores the utility of BaRE in an operational setting involving the prediction of streamflow for an "uncalibrated" watershed using the SAC-SMA model developed by *Burnash et al.* [1973]. In both cases, we predict a scalar-valued output (i.e., $m = 1$), assume a Gaussian error model (i.e., $\beta = 0$), and use a uniform distribution on θ as the prior distribution $p(\theta)$.

4.1. Study Watershed and Data

The hydrologic data used in the case studies consist of ~11 years (July 28, 1952, to September 30, 1962) of observation time series from the Leaf River basin. This humid 1944 km^2 watershed, located in southern Mississippi, has been investigated intensively [e.g., *Sorooshian et al.*, 1983; *Brazil*, 1988]. The data, obtained from the Hydrologic Research Laboratory, consist of mean areal precipitation (mm per 6 hours), potential evapotranspiration (mm/d), and streamflow (m^3/s). The mean annual precipitation for the entire period, excluding the 65-day spin-up period in water year (WY) 1952, is 1324 mm, and the mean annual runoff is 27.13 m^3/s. Data from WY 1953 are used extensively in this study. The mean annual precipitation for that period is also 1324 mm, but the mean runoff is 33.60 m^3/s, more than 10% higher than the 11-year average.

4.2. Case Study I: Nash-Cascade Model

In this section, we investigate the general characteristics of the BaRE method by means of a synthetic data study using the two-parameter Nash-cascade model, a very simple unit hydrograph model. The use of synthetic output data (generated using the observed precipitation and an assumed true parameter set) allows an assessment of the prediction and parameter estimation potential of the Bayesian methodology under the controlled conditions of no errors in model structure and known properties of the data measurement error.

The Nash-cascade model is a channel routing scheme that is based on a series of linear reservoirs [*Nash*, 1960]. Conceptually, the inflow is successively routed through n reservoirs that all have the same recession coefficient k. Mathematically, this response function can be written as

$$h(t) = \frac{1}{k\Gamma(n)} \left(\frac{t}{k}\right)^{n-1} \exp\left\{-\frac{t}{k}\right\}, \qquad n > 0, \qquad k > 0, \tag{29}$$

where n represents the number of reservoirs (the continuous mathematical formulation allows n to be noninteger), k represents the average travel time through each reservoir, t denotes the elapsed time from the time of impulse, and $\Gamma(\ )$ is the gamma function. This scheme can be used as a very simple rainfall-runoff model, where n and k describe the unit hydrograph of the watershed of interest. Because n and k are unknown conceptual properties of the simulated watershed, they must be estimated via calibration.

Synthetic daily output data (in cm^3/s) were generated for the period June 28, 1952, to September 30, 1953, by driving the Nash-cascade model with mean areal rainfall and fixed values for the parameters ($n = 2.0$ (dimensionless) and $k = 25.0$ (1/6 hours). The synthetic true output time series was perturbed by imposing normally distributed noise, with mean zero and a variance of 20% of the mean of the true flows, to generate synthetic "measured" streamflow data.

A variety of experiments were performed to assess the effectiveness of the BaRE methodology in terms of its ability to generate accurate and precise streamflow predictions and to locate the region of the true parameter values. We were also concerned with algorithm efficiency, particularly the amount of data required by the algorithm to converge to the known answers. To explore the sensitivity of the algorithm to the way in which the feasible parameter space is sampled, we made several runs using three different sampling strategies. The feasible space was defined by allowing n to vary from 1.0 to 3.0 and k to vary from 2.0 to 48.0 days. The initial guess for the variance σ_0^2 was set to 5.8 (equal to the variance of the streamflow). To reduce sensitivity to initialization, a 65-day warmup period was used during which no updating of the density functions was performed. After that, the observed data were assumed to be available at each time step. The results obtained for each sampling strategy are discussed below. In the interest of brevity, graphical results are presented only for strategy 2.

4.2.1. Strategy 1. In this experiment the parameters n and k were sampled over the feasible space using a rectangular grid of 21 by 21 points (total 441 points) that contains the true parameter values at the midpoint of the grid. As expected, the BaRE algorithm pinpointed the true parameter values for both the "perfect data" case and the 20% measurement error case. For the perfect data case the algorithm exactly reproduced the true hydrograph, and the prediction uncertainty reduced to zero after only three time steps. For the 20% measurement error case the parameter uncertainty (and hence the associated output prediction uncertainty) became insignificant after ~75 days. In addition, the 95% Bayesian confidence intervals associated with the output measurement reached a constant width that included ~90% of the measured streamflow values, slightly less than the expected value of 95%.

4.2.2. Strategy 2. In this experiment the parameters n and k were sampled over the feasible space using a rectangular grid of 20 by 20 points such that the sampled set does not contain the true parameter values (a more realistic situation). In this case, the BaRE algorithm resulted in a slower but steady convergence of the HPD region into the vicinity of the true parameter values. Figure 1 shows the evolution of the HPD region of the posterior parameter probability density (in the form of one-dimensional projections, with darker shading indicating higher probability) for the case of 20% measurement error. Notice that it took ~100 days to detect the most likely parameter set (a grid point just next to the true parameter values, indicated by an asterisk on the plots) and ~125 days for the algorithm to assign a probability >95% to the same parameter set. Figure 2 shows three snapshots, at time steps of 50, 75, and 100 days, of the posterior joint density of the two model parameters. Figure 2a shows that by day 50 the BaRE methodology has begun to assign higher probabilities to ~25 points in the region having higher k and lower n than the true values ($k_{true} = 24$ days, $n_{true} = 2.0$). By day 75 the algorithm has concentrated on 12 points arranged along an arc, indicating interaction between the parameters, but has selected two possibilities for the most likely points (Figure 2b). By day 100, one of the four grid points surrounding the true parameter values has been assigned the highest probability, although there are still two other possible competitors (Figure 2c).

Figure 3 shows how the BaRE algorithm translates the parameter uncertainty into model predictions of the true and measured streamflows. Figure 3a shows the predicted and observed streamflows, and Figure 3b shows the residuals and prediction uncertainties measured in terms of the differences from the true streamflow values. The dark shaded region, representing the 95% Bayesian confidence interval for the prediction of the true streamflows, is associated with the current uncertainty in the parameter estimates. This region is seen to be relatively wide for the first 140 days, reflecting the imprecision in the parameter estimates but collapses to a line when the algorithm settles on a best parameter set. Notice that the results indicate a small tendency toward positive bias, reflecting the fact that the true parameters are not contained in the feasible sample. The light shaded region represents the 95% Bayesian confidence interval for the prediction of the streamflow measurement (i.e., the total prediction uncertainty associated with parameter, structure, and measurement errors). In contrast to the dark shaded region, this confidence interval tends toward a constant width centered on the "most likely" prediction and is found to bracket ~90% of the observed data. In addition, the unreasonably large initial value assigned to σ_0^2 is reflected in the exponentially decreasing width for the uncertainty region during the first 30 days. In general, the impact of the choice of σ_0^2 decays rapidly and does not substantially affect the overall results.

4.2.3. Strategy 3. In general, the method of regular sampling from a rectangular grid is only useful for models having a small number of parameters (one to four) because the sample size in a grid-sampling scheme grows exponentially with parameter dimension. A more practical approach would involve random sampling of a predetermined number of points from the feasible space in such a way as to uniformly sample all the parameter regions of interest. Therefore, in this experiment the parameters n and k were sampled randomly over the feasible space using a uniform distribution for a total of 441 points. (Note that the sample did not include the true parameter set.) The number of points was selected to correspond to the numbers of points in the earlier strategies. The performance of this strategy was comparable to the grid-based approach, with the algorithm converging to one of the parameter sets close to the true values after ~120 days.

To test the sensitivity of the algorithm to sample size, a second run was made using a sample of only 25 points distributed uniformly over the feasible space. In this case, the algorithm quickly selected a single best parameter set at ~50 days but then jumped back and forth several times between this and two other parameter sets before finally selecting one of them at ~220 days. The reason for this "jumping" behavior is that unlike the earlier grid-sampling strategy, the relatively sparse

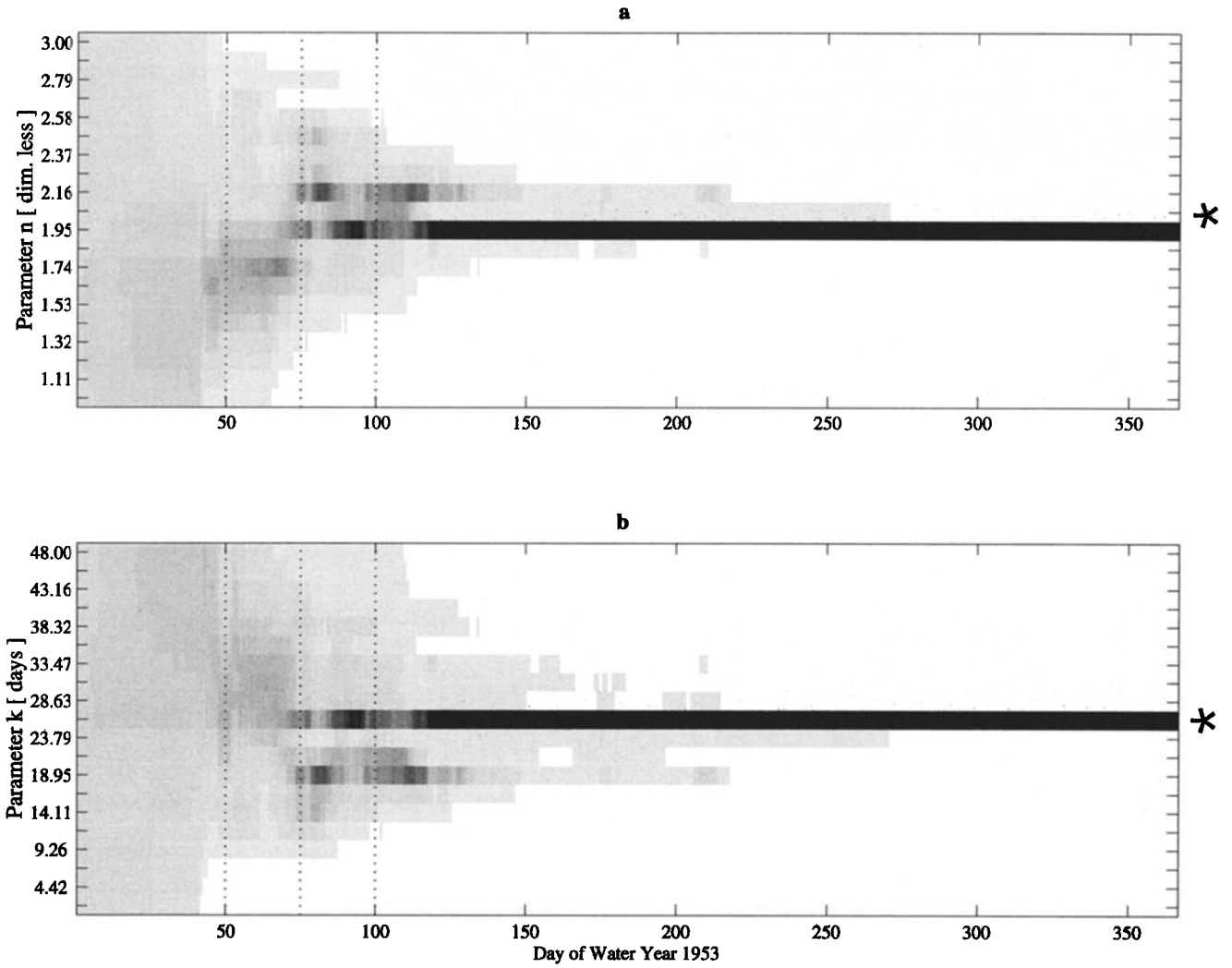

Figure 1. Evolution of the 99.99% parameter high probability density region for the Nash-cascade model (strategy 2, 20% measurement error) for (a) n and (b) k. Darker shading indicates higher probabilities. The asterisk denotes the "true" parameter values.

and irregular sampling of the feasible space did not (in this case) result in a very good potential candidate. This was also reflected in poorer overall simulation of the streamflow hydrograph and wider prediction uncertainty bounds (for details, see *Thiemann* [1999]).

4.3. Case Study II: Sacramento Model

The SAC-SMA model (see Figure 4) is one of the components of the National Weather Service River Forecast System. This conceptual rainfall-runoff model, first introduced by *Bur-*

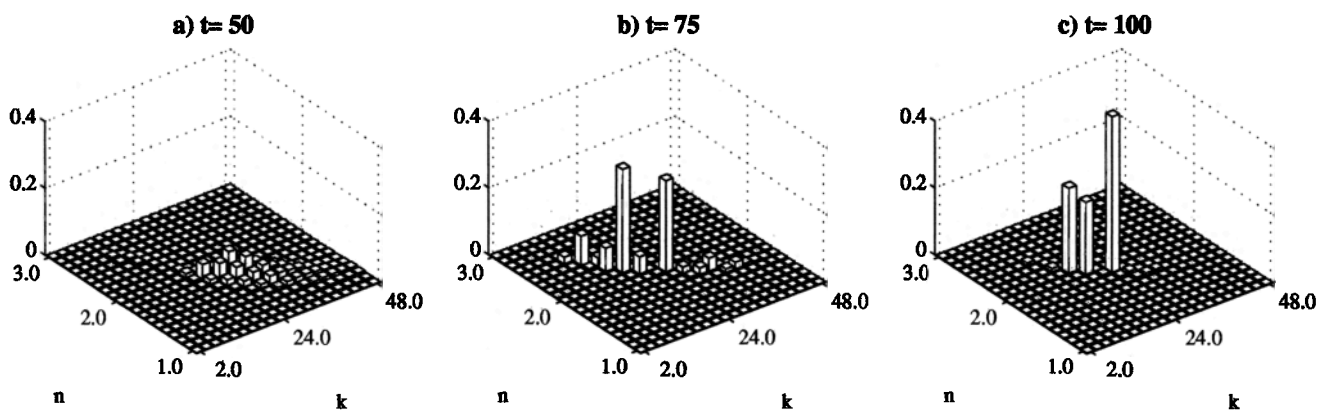

Figure 2. Snapshots of the posterior joint density of the two Nash-cascade model parameters n and k at time steps (a) 50, (b) 75, and (c) 100 days.

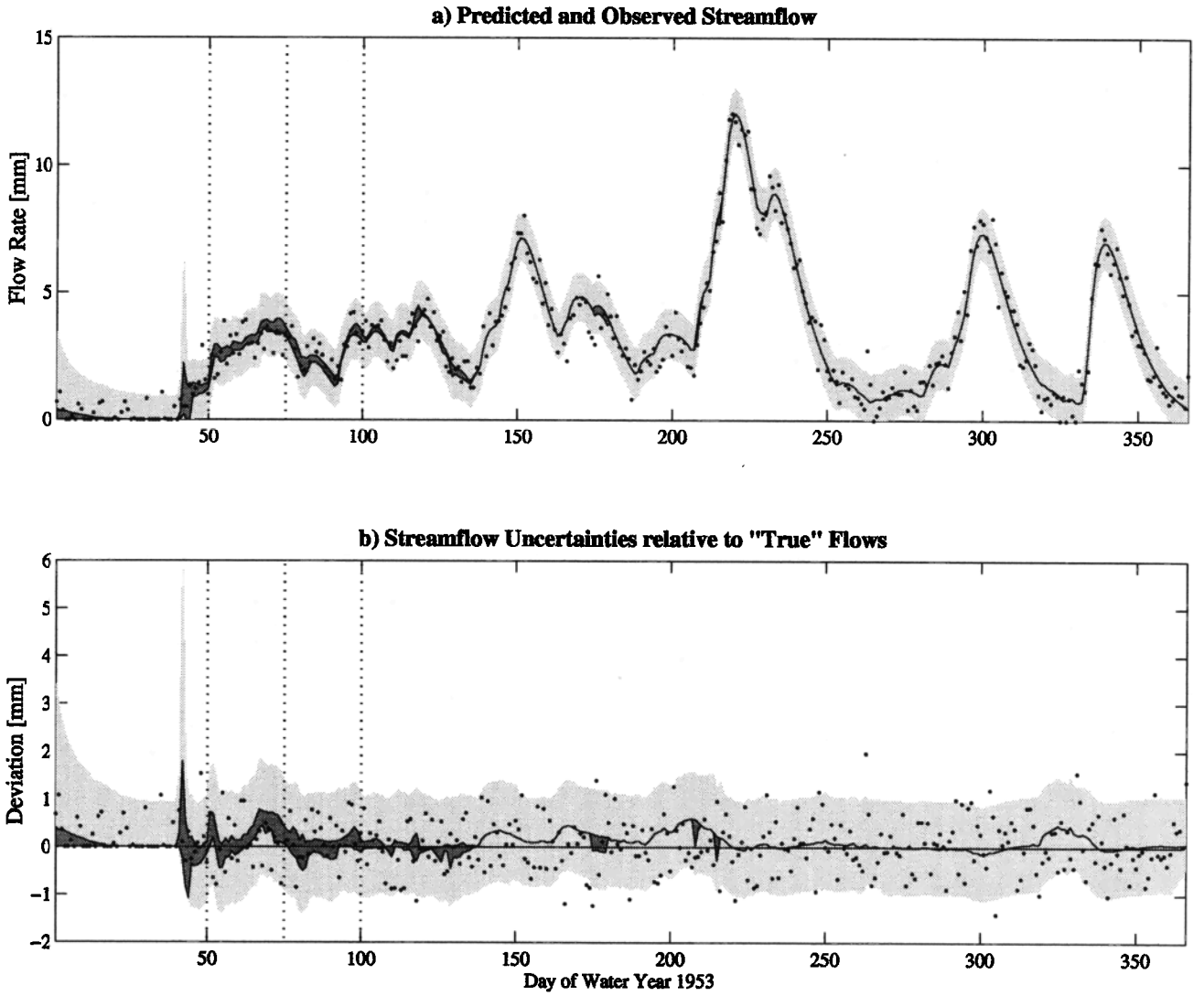

Figure 3. Probabilistic streamflow predictions made using the Nash-cascade model (strategy 2, 20% measurement error). Solid dots denote the "measured" streamflow, the dark shaded region indicates the 95% confidence intervals for prediction of the "true" streamflow, and the light shaded region indicates the 95% confidence intervals for prediction of the measured streamflow. (a) streamflow predictions and (b) streamflow uncertainties relative to "true" flows.

nash et al. [1973], has been used extensively for parameter estimation and streamflow forecasting studies [e.g., *Sorooshian and Gupta*, 1983; *Brazil*, 1988; *Duan et al.*, 1994; *Gupta et al.*, 1998]. This case study illustrates the usefulness of the BaRE methodology to hydrologists who wish to use the SAC-SMA model for flow forecasting on a watershed that has not yet been calibrated and for which data monitoring has only recently been initiated. Hydrologists must therefore make the best use of the limited amounts of available data, crude parameter estimates (ranges) based on nearby watersheds, and new gauge data as they become available.

In keeping with previous studies [e.g., *Brazil*, 1988; *Yapo et al.*, 1996], 13 model parameters were assumed to be unknown and to have the uncertainty ranges defined in Table 1. The feasible parameter space defined by these ranges was uniformly sampled at 10,000 randomly selected locations and was assigned a uniform prior probability density function. Further, we assumed that the output errors have a heteroscedastic (nonconstant) variance that is related to flow level and which can be stabilized by the Box-Cox transformation $z = (y^{\lambda} - 1)/\lambda$ with $\lambda = 0.5$ [see, e.g., *Sorooshian and Dracup*, 1980]. The model simulations were assumed to begin on July 28, 1953, with measured streamflow data only becoming available on October 1, 1953, at which point, operational forecasting and model updating using the BaRE methodology was initiated. Hence the initial 65 days represent a spin-up period to help minimize model initialization errors. From October 1 onward, BaRE was used to generate one-day-ahead streamflow predictions using the current parameter probability distribution and to update the parameter probability distribution daily as soon as a new streamflow measurement became available. The initial guess for the variance σ_0^2 was set to 10.0, a value selected to be twice the variance of the streamflows for illustrative purposes only. In general, σ_0^2 should normally be selected to reflect the anticipated size of the prediction uncertainty, which will typically be much smaller than the variance of the streamflows.

The streamflow prediction results for WY 1953 are shown in

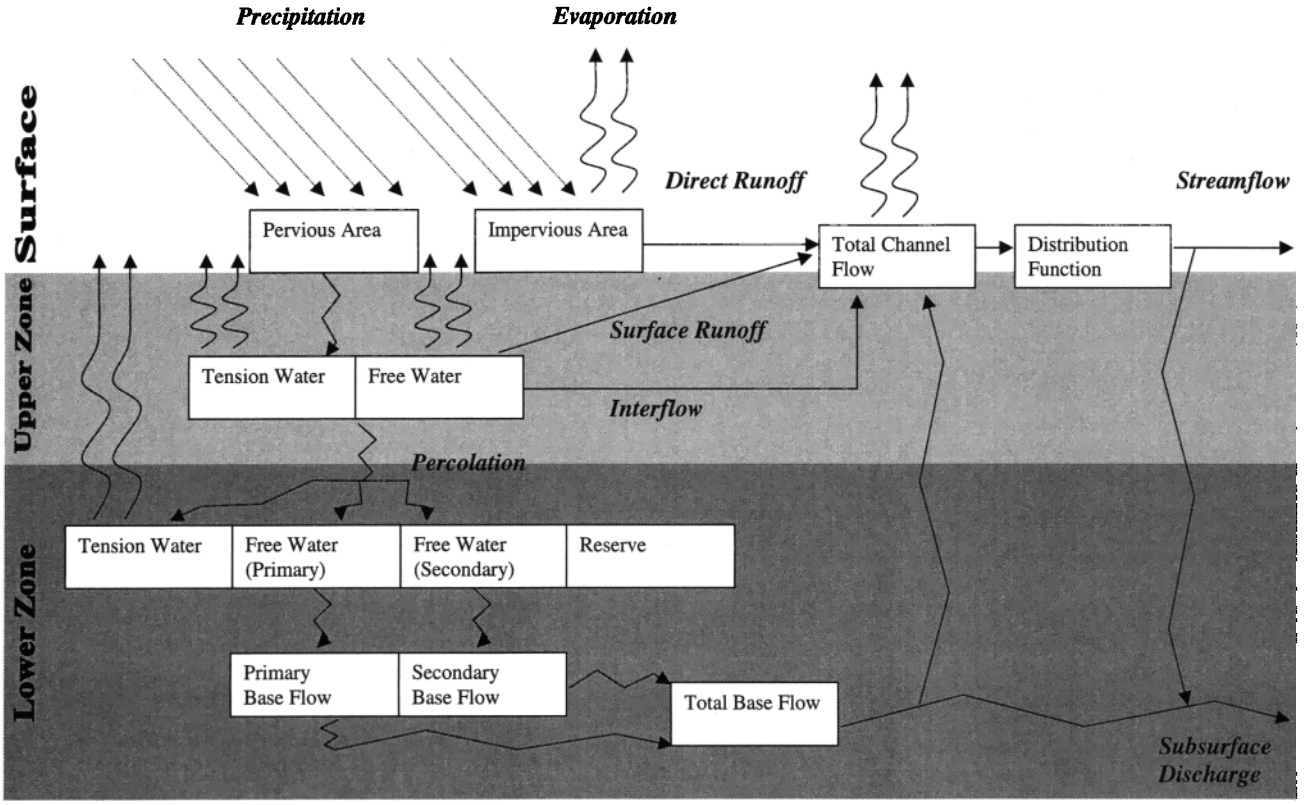

Figure 4. Conceptual flow diagram of the SAC-SMA model.

Figure 5. Figures 5a and 5b show the results in the original and in the transformed output spaces, respectively, while Figure 5c shows the residuals (for the transformed outputs) measured in terms of the differences from the maximum likelihood streamflow predictions. Notice that the model predictions appear to track the observations fairly well. The dark shaded region, representing the 95% Bayesian confidence interval for the prediction of the true streamflows, narrows very quickly in the first 35 days but continues to widen and narrow intermittently (until about day 200) each time the algorithm is forced to deal with larger than expected prediction errors. Thereafter, the confidence region collapses to a line, reflecting the fact that the probability mass associated with the most likely parameter set has become large enough not to be further disturbed during the course of the run. Further, the residuals (in transformed output space) during those first 200 days are relatively uncorrelated but thereafter show a tendency toward systematic positive bias (predictions larger than observations). These results

Table 1. Comparison of Parameter Values Estimated by *Brazil* [1988] (Interactive Multilevel Calibration), SCE-UA, and BaRE

	Characteristics				Models		
Parameter	Unit	Lower Bound	Upper Bound	Fixed Value	Brazil	SCE-UA	BaRE
UZTWM	mm	1.0000	150.0000		9.000	14.089	33.610
UZFWM	mm	1.0000	150.0000		39.800	63.825	76.120
UZK		0.1000	0.5000		0.200	0.100	0.332
PCTIM		0.0000	0.1000		0.003	0.000	0.016
ADIMP		0.0000	0.4000		0.250	0.363	0.266
RIVA				0.0000			
ZPERC		1.0000	250.0000		250.000	249.972	117.300
REXP		1.0000	5.0000		4.270	2.459	4.948
LZTWM	mm	1.0000	1000.0000		240.000	237.779	235.600
LZFSM	mm	1.0000	1000.0000		40.000	3.191	131.900
LZFPM	mm	1.0000	1000.0000		120.000	99.826	123.500
LZSK		0.0100	0.2500		0.200	0.019	0.089
LZPK		0.0001	0.0250		0.006	0.021	0.015
PFREE		0.0000	0.6000		0.024	0.001	0.146
RESERV				0.3000			
SIDE				0.0000			
DRMS[a]					20.3	18.2	21.8

[a]Daily root-mean-square error of prediction for the 11-year period WY 1953–1963.

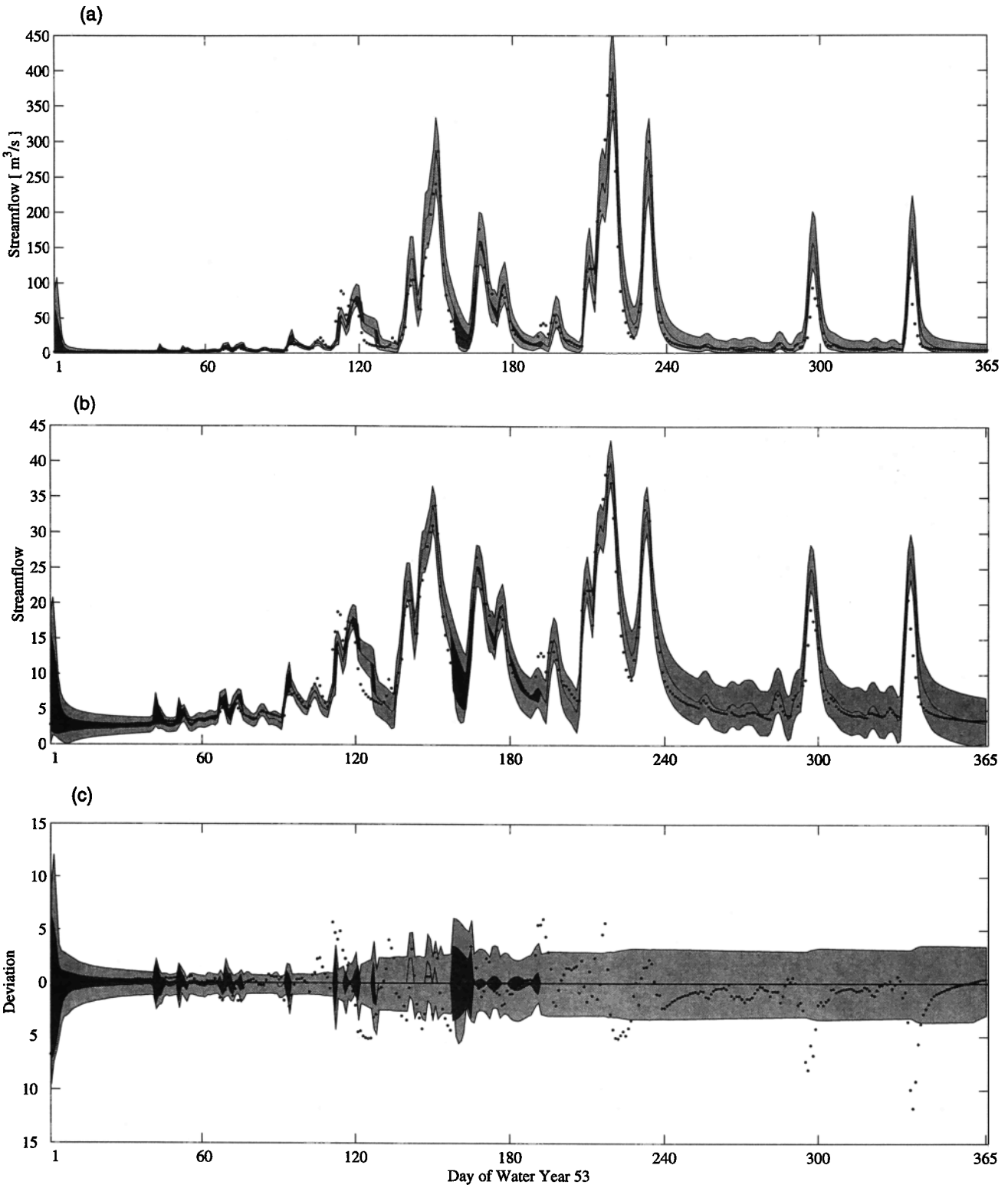

Figure 5. Probabilistic streamflow predictions made using the SAC-SMA model for the Leaf River basin, Mississippi (WY 1953). Solid dots denote the measured streamflow, the dark shaded region indicates the 95% confidence intervals for prediction of the "true" streamflow, and the light shaded region indicates the 95% confidence intervals for prediction of the measured streamflow: (a) streamflow predictions, (b) streamflow predictions in transformed output space, and (c) streamflow uncertainties relative to the most probable forecast in the transformed output space.

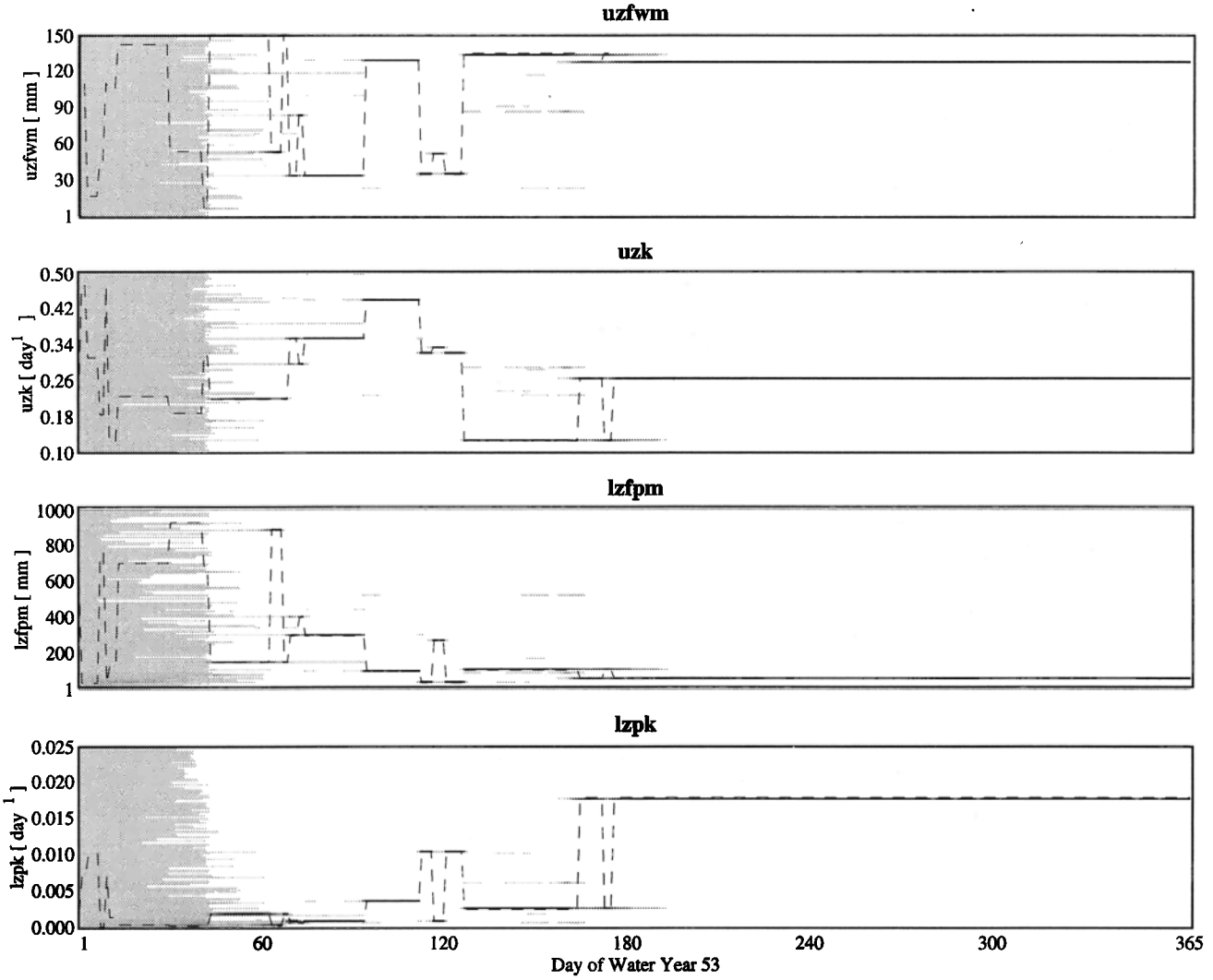

Figure 6. Evolution of the 99.99% parameter high probability density region for the SAC-SMA parameters uzfwm, uzk, lzfpm, and lzpk for WY 1953. Shading indicates higher probabilities, and the dashed line indicates the most probable parameter value.

suggest that too much confidence may have been assigned to the model, a point we will return to in section 5. In contrast, the width of the light shaded region, representing the 95% Bayesian confidence interval for the prediction of the streamflow measurement (in transformed output space; see Figure 5c), decreases to a relatively small value during the first 100 relatively dry days, thereafter doubling over the next 3 months of significant rainfall activity. This variation in estimated prediction uncertainty suggests that a better choice for the error model may be possible.

The evolution of the parameter estimates is illustrated in Figures 6a–6d for four of the model parameters upper zone free water capacity m (uzfwm), upper zone free water withdrawal rate k (uzk), lower zone primary free water capacity m (lzfpm), and lower zone primary free water withdrawal rate k (lzpk). This plot shows a projection onto each parameter axis of the parameter values for the points belonging to the 99.99% HPD region at each time step. In the beginning, due to the uniform prior probability, all the points appear on the plot (see the first few time steps). As the streamflow data are processed, the number of points belonging to the updated HPD region decreases rapidly, leaving only a few competing parameter sets. However, the competing parameter sets appear to be scattered at different locations within the feasible parameter space, indicating that different parameterizations of the SAC-SMA model can give rise to a similar ability to predict the outputs for the limited amount of data processed so far. In particular, the HPD for the lower soil zone recession constant lzpk first tends toward the lower portion of the feasible space but later shifts toward a much higher value.

Also marked in Figure 6 is the location of the most probable (maximum likelihood) parameter set at each time step (dashed line). In the beginning, the most probable value jumps back and forth from one to another of the competing parameter sets but finally settles down to a best value at about day 200. It is interesting to note that the major jumps coincide with some of the large storm events. The similarity of this jumping phenomenon to that observed in strategy 3 of the synthetic case study suggests that a possible cause is insufficient density of sampling in the region of good parameter values. However, one should

also consider the possibility that the assumed models (i.e., hydrologic model, error model, and transformation model) do not adequately represent the observed input-output process.

Finally, Table 1 shows a comparison of the 11-year forecast performance and final parameter estimates obtained using BaRE in on-line mode with two conventional off-line batch calibration runs, one reported by *Brazil* [1988] and the other performed by us. This comparison is possible because all three studies involve calibration of the SAC-SMA model to the Leaf River basin using the 11-year period WY 1953–1963. *Brazil* [1988] used a three-stage interactive multilevel calibration procedure that combined manual and automatic methods, while we used the SCE-UA global optimization procedure developed by *Duan et al.* [1994]. Note that while the batch methods have the advantage of being able to employ sophisticated search procedures to actively refine the location of the parameter estimates, the current version of the BaRE procedure is limited to selecting from a fixed set of randomly specified points distributed rather coarsely throughout the feasible space. In spite of this, the BaRE method provided forecasts with an overall RMSE prediction error that is only slightly larger than that given by the *Brazil* [1988] results. Furthermore, the parameter differences are not very large (compared to the initial uncertainty ranges), with the exception of uzk, maximum percolation rate coefficient (zperc) (dimensionless), and percolation equation exponent (rexp) (dimensionless), which are known to be generally less sensitive. It is interesting to note that the BaRE method converged to its final parameter estimates after processing only ~1.5 years of data. This is impressive, considering that at least 11 years of data are generally considered necessary for obtaining reliable parameter estimates via conventional batch calibration.

5. Summary and Discussion

This paper has discussed a Bayesian probability-based approach to simultaneous parameter estimation and output uncertainty prediction via on-line recursive processing of time series data as they become available. A practical algorithm, called the Bayesian Recursive Estimation (BaRE) method, has been developed and tested. BaRE differs from conventional calibration and prediction methods in that it employs a recursive scheme for tracking the conditional probabilities associated with several competing parameter sets (models) in an on-line mode instead of searching for a single best solution in an off-line mode. The parameter probabilities are used to compute probabilistic predictions of the desired output variables. Probability updating, via Bayes theorem, facilitates the assimilation of new data as they become available.

The most obvious practical advantage of the BaRE approach is that it enables hydrologists to generate consistent model predictions, along with estimates of the uncertainty in those predictions, even if historical hydrologic data are not available for model calibration using conventional methods. While the initial predictions (and parameter estimates) may be crude, their quality can be refined recursively with time as data become available. The approach is therefore directly applicable when a watershed has only recently been gauged. However, this approach could also be extremely useful where the time and expertise required for conventional calibration are not available. In such situations the parameter estimates provided by BaRE could help to provide guidance to later model calibration efforts. Furthermore, we believe that the BaRE procedure can be a very efficient and valuable tool for the refinement of parameters defined through regionalization, i.e., the transfer of parameters from a calibrated watershed to another watershed with similar hydrologic characteristics. In this case, the regionalized parameter values can be used to determine parameter ranges and prior distribution. This distribution can then be altered successively by BaRE to reflect the subtle differences between the calibrated and regionalized watershed.

The two hydrologic case studies presented in this paper serve to illustrate the effectiveness and efficiency of the BaRE methodology, as well as some weaknesses of the current implementation. The first (synthetic data) study used a simple unit hydrograph model to show that the BaRE algorithm does indeed provide consistent one-step-ahead probabilistic predictions of the true and measured streamflow values. At the same time, the algorithm is able to quickly and precisely locate the region of the known true parameter values, even when the data contain significant amounts of noise. The second (real data) case study demonstrated the applicability of BaRE in the context of the SAC-SMA model used by the U.S. National Weather Service for operational streamflow forecasting. Although the SAC-SMA model cannot be considered to be a perfect representation of the watershed process for the Leaf River basin, the BaRE method produced good one-step-ahead flow predictions with reasonable uncertainty estimates, with an accuracy that compares well with traditional batch calibration methods. In both cases the algorithm was relatively insensitive to the initial estimate σ_0^2 for variance of the errors (for practical applications, we suggest setting σ_0^2 at ~25–50% of the estimated long-term variance of the streamflows). Furthermore, it was somewhat surprising that BaRE converged to its final most likely parameter estimates after processing only a relatively small amount of data. The consistency of this finding needs to be evaluated through further studies.

Our experiences with the current version of the BaRE algorithm suggest that there are several ways in which the implementation can be improved. The first is related to the phenomenon where the current most likely parameter set jumps large distances in the feasible parameter space. While this may be a natural consequence of parameter insensitivity, interaction, or nonuniqueness of the model structure, we suspect that a probable cause is a combination of uneven and insufficient density of sampling in the region of good parameter values. It is interesting to note that after all the available data have been processed, the final most likely BaRE parameter estimate is theoretically equivalent to the value that would have been obtained by conventional batch calibration using pure random search. However, state-of-the-art batch calibration methods iteratively refine the search to focus on the most promising region of the parameter space. Thus we are exploring ways to implement progressive resampling in order to concentrate the samples in the current HPD region while terminating computations in the nonproductive portions of the parameter space. At the same time, special procedures such as Latin hypercube sampling [e.g., *Ye*, 1998] can be implemented to control sample evenness.

The second area of improvement is related to the selection of the output transformation and error models. The present algorithm assumes that the correct values for β and λ are known and that the error variance in the transformed output space is constant. However, in the case of the SAC-SMA study the 95% Bayesian confidence interval for the prediction of the streamflow measurement varied substantially, suggesting the

need to investigate better choices for the transformation and error model parameters. We recognize, therefore, the need for objective procedures to guide the selection of the appropriate transformation and error models.

Finally, the SAC-SMA case study also revealed that the model predictions tended to be unbiased during the early portion of the run when the parameter HPD region had not yet collapsed to a single point. However, during the later period, when the algorithm had decided on a single best parameter set, the residuals showed a tendency toward systematic positive bias. We speculate that the performance of the BaRE methodology might be improved by preventing the parameter HPD region from collapsing to a single point. We believe that the algorithm might overvalue the information in the most recent data, and we are currently testing different ways in which this could be prevented, such as the incorporation of a "forgetting factor" into the parameter probability updating rule. Such a strategy may also have the additional benefit of allowing the algorithm to track slowly varying changes in the hydrologic behavior of the watershed (e.g., due to deforestation). Our findings related to these potential algorithmic improvements will be reported in future papers.

Acknowledgments. The ideas for this research arose out of discussions within the Surface Water Modeling and Calibration Group at the Department of Hydrology and Water Resources. We want to acknowledge Luis Bastidas, Douglas Boyle, Terri Hogue, Kuo-lin Hsu, and Tom Meixner for helpful input and comments. This material is based upon work supported in part by the National Weather Service (grants NA87WHO581, NA87WHO582, and NA86GPO324), the National Aeronautics and Space Administration (grant NAG5-3640 and NASA-EOS VPR matching funds), and by SAHRA (Sustainability of Semi-Arid Hydrology and Riparian Areas) under the STC program of the National Science Foundation, agreement EAR-9876800. Computational support was provided by Jim Broermann and Dan Braithwaite, and Corrie Thies and Sharrin Wilson proofread and edited the manuscript.

References

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490, 1999.

Beven, K. J., Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*, 41–51, 1993.

Beven, K. J., and A. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298, 1992.

Binley, A. M., and K. J. Beven, Physically-based modelling of catchment hydrology: A likelihood approach to reducing predictive uncertainty, in *Computer Modelling in the Environmental Sciences*, edited by D. J. Farmer and M. J. Rycroft, pp. 75–88, Clarendon, Oxford, England, 1991.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley-Longman, Reading, Mass., 1973.

Brazil, L. E., Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. dissertation, Colo. State Univ., Fort Collins, 1988.

Burnash, R. J. E., R. L. Ferral, and R. A. McGuire, A generalized streamflow simulation system, report, Joint Fed.-State River Forecast Cent., Sacramento, Calif., 1973.

Duan, Q., S. Sorooshian, and V. K. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, 1992.

Duan, Q., V. K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*(3), 501–521, 1993.

Duan, Q., S. Sorooshian, and V. K. Gupta, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284, 1994.

Fisher, A., *The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods*, Macmillan, Old Tappan, N.J., 1922.

Franks, S. W., and K. J. Beven, Bayesian estimation of uncertainty in land surface-atmosphere flux predictions, *J. Geophys. Res.*, *102*(D20), 23,991–23,999, 1997.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot, On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.*, *34*(4), 787–797, 1998.

Freer, J., K. J. Beven, and B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, 1996.

Gupta, V. K., S. Sorooshian, and P. O. Yapo, Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763, 1998.

Gupta, V. K., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503, 1999.

Jones, D. A., Statistical analysis of empirical models fitted by optimization, *Biometrika*, *70*, 67–88, 1983.

Kuczera, G., On validity of first-order prediction limits for conceptual hydrological models, *J. Hydrol.*, *103*, 229–247, 1988.

Kuczera, G., G. P. Raper, N. S. Brah, and M. D. Jayasuriya, Modelling yield changes after strip thinning in a mountain ash catchment: An exercise in catchment model validation, *J. Hydrol.*, *150*, 433–457, 1993.

Mroczkowski, M., G. P. Raper, and G. Kuczera, The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, *33*(10), 2325–2335, 1997.

Nash, J. E., A unit hydrograph study with particular reference to British catchments, *Proc. Inst. Civ. Eng.*, *17*, 249–282, 1960.

Romanowicz, R., K. J. Beven, and J. A. Tawn, Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach, in *Statistics for the Environment*, vol. 2. *Water-Related Issues*, edited by V. Barnett and K. F. Turkman, pp. 297–317, John Wiley, New York, 1994.

Sorooshian, S., Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors: The noninformative data case, *J. Hydrol.*, *52*, 127–138, 1981.

Sorooshian, S., and J. A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442, 1980.

Sorooshian, S., and V. K. Gupta, Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268, 1983.

Sorooshian, S., V. K. Gupta, and J. L. Fulton, Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility, *Water Resour. Res.*, *19*(1), 251–259, 1983.

Sorooshian, S., Q. Duan, and V. K. Gupta, Calibration of rainfall-runoff models: Application of global optimization to the soil moisture accounting model, *Water Resour. Res.*, *29*(4), 1185–1194, 1993.

Spear, R. C., and G. M. Hornberger, Eutrophication in Peel Inlet, 2, Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*, 43–49, 1980.

Spear, R. C., T. M. Grieb, and N. Shang, Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, *30*(11), 3159–3169, 1994.

Thiemann, M., Uncertainty estimation of hydrological models using Bayesian inference methods, M.S. thesis, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1999.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, *181*, 23–48, 1996.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multiobjective global optimization for hydrologic models, *J. Hydrol.*, *204*, 83–97, 1998.

Ye, K. Q., Orthogonal column Latin hypercubes and their application in computer experiments, *J. Am. Stat. Assoc.*, *93*(444), 1430–1439, 1998.

H. Gupta and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Harshbarger Bldg., Room 122, Tucson, AZ 85721. (hoshin@hwr.arizona.edu; soroosh@hwr.arizona.edu)

M. Thiemann, Riverside Technology, Inc., 2290 East Prospect Road, Suite 1, Fort Collins, CO 80525. (mt@riverside.com)

M. Trosset, Department of Mathematics, College of William and Mary, Williamsburg, VA 23187. (trosset@math.wm.edu)

(Received September 29, 2000; revised December 14, 2000; accepted December 14, 2000.)